Data Collection through Vehicular Sensor Networks by using TCDGP

P Radha Krishna Reddy[1]
CSE & JNTU
Anantapur

P Joshna
SE & Sri Vidyanikethan
Tirupati

G.Sireesha
CSE & JNTU
Hyderabad

A.Thirupathaiah
CSE & JNTU
Kakinada

***Abstract*— Now a days Many car manufacturers are planning to deploy wireless connectivity equipment in vehicles to enable communications with "roadside base station" and also between different vehicles, for the purposes of safety, driving assistance, and entertainment. The distinct feature of vehicle is that they are highly mobile, with speed up to 30 m/s, though their mobility patterns are more foreseeable than those of nodes in Mobile Ad-hoc Networks (MANET) due to the conditions are forced by road, speed limits, and commuting habits. Therefore, these networks require specific solutions and identify a novel research area, i.e., Vehicular Ad-hoc Networks (VANET). In this paper, mainly we focus on a particular VSN architecture, where the ad hoc network is operated by a telecommunication/service provider to combine non-valuable individual sensed data and extract from them effective feedbacks about the situation of the road in a geographical area. In operated VSNs, providers tend to reduce the traffic load on their network, using the free-frequency communication medium (IEEE 802.11p). To do so, we propose TCDGP (Tree based Clustered Data Gathering Protocol), a cross layer protocol based on efficient historical data collection, aggregation and dissemination mechanisms. We analyse the performances of our solution using a simulation environment and realistic mobility models. We are showing the feasibility of such a solution.**

***Keywords*— VSN, VANET, dissemination, ITS, data collection, data aggregation, hybrid architecture, operated network.**

1. INTRODUCTION

Over the past decade the nature of wireless communications has evolved rapidly. The introduction of third generation and Wireless LAN technologies and the recent technical implementation of WiMax have helped to realize the vision of ubiquitous connectivity. Currently, much research effort is focusing on exploiting this "always-on" feature for use in Transportation Systems. The primary objective of ITS (Intelligent Transportation Systems) is to improve traffic safety, efficiency, and travelling comfort.

Vehicular Sensor Networks are made on the top of VANET by preparing vehicles with onboard sensing devices. Here, sensors grouped for not only safety-related information, but also more complex multimedia data like videos. Compare with the traditional sensor networks, VSNs are not subject to major memory, processing, storage, and energy limitations (Giovanni pau). However, distinctive scale of a VSN about wide geographic areas, the volume of generated data like live streaming, and mobility of vehicles make it unrealizable to adopt traditional sensor network solutions where sensed data tends to be in a particular way delivered to sinks using data-centric protocols like Directed Diffusion. Further, the flexibility of sensor nodes makes it less efficient to use mobile agents, or static sensor networks, it picks data from sensors when in near by, buffer it, and drop off the data to wired access points.

Besides DSRC, we can utilize cellular communications (2/3G) via Smartphone's. The Smart phone's are modelled with various sensors such as GPS, camera, audio, and video, and support various communications means such as 2/3G, Wi-Fi, and Bluetooth. Bluetooth enables us to connect other external sensors via a wireless data acquisition board. The importance of 2/3G connection is that it gives always-on Internet connection, which makes data access and retrieval amenable.

In this system, we focused on the main component is the Intelligent Transportation Systems, which is the communication between vehicles. Indeed, many car manufacturers are deploying wireless connectivity mechanism in their vehicles to enable communication between vehicles. Vehicular Sensor Networks (VSNs) can be built on top of these vehicular networks by equipping vehicles with onboard sensing devices (Gurupreet Singh). In such case, sensors can gather a set of information like video data, speed, localization, acceleration, temperature, seat occupation, etc. Compared with traditional sensor networks, this is the one recently emerged sensor network is not restricted by the power supply and the storage space. However, the typical scale of a VSN over wide areas, the volume of generated data like streaming video, and mobility of vehicles make it infeasible to adopt

traditional sensor network solutions where as sensed data tends to be systematically delivered to sinks using data-centric protocols such as Directed Diffusion [1].

An effective and efficient architecture is used for data collection and data exchange is more important in this architecture. This work deals with the system framework that consists of mobile vehicular sensors and road-side-units operated by an operator or a service provider (WiMax access point, 2.5/3G base station). Road-side-units are distributed over the road for collecting data from mobile vehicular sensors passing by it. While mobile sensors on vehicles senses and send that information to the road side units.

In this study, the VSN will be used by the owner of the infrastructure wireless networks like Internet provider to gather "useless" individual information from each vehicle and to aggregate them inside an ad hoc wireless network using free frequency, to get a global view of the state of the road in a geographical area at a specific time, or to use these information as a database for a posterior treatment shown infigure1.

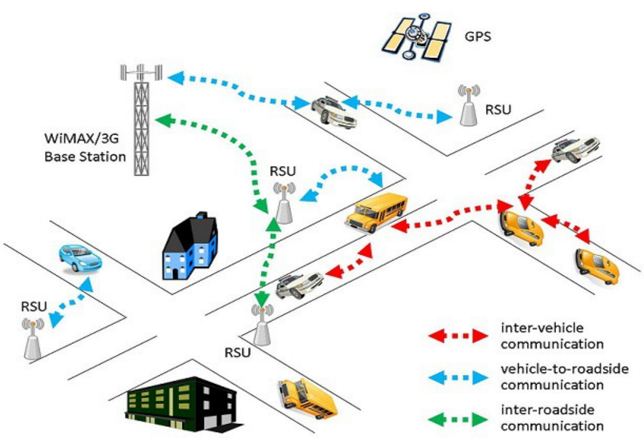

Figure 1 – Operated vehicular sensors network

In this study, the VSN will collect individual information from each and every vehicle and aggregate it inside the ad hoc wireless network. The aggregated information will be sent to the road side operator via a non-free frequency (WiMax or 2.5/3G). In fact, these sensors may generate enormous amounts of censored data and there is a need of collection, storing, and retrieving. The objective in this architecture is an operator/service provider is to reduce the use of its high-cost links. To do so, we present TCDGP (Tree based Clustered Data Gathering Protocol): a cross-layered protocol based on hierarchical and geographical data gathering, aggregation and dissemination by using tree like a structure. The goal of TCDGP is to gather data from all nodes in the vehicular ad hoc networks in order to offer different kind of ITS services.

- A real-time traffic information service, by gathering all node's positions and velocities, [2]
- A geographical localization service for customers who want to follow their vehicles mobility (fleet management),
- A parking lots availability service, by detecting empty spaces in parking lots, [3]
- Warnings messages in a specific area, when an unusual event happens (a sudden speed decrease of several vehicles, for example), [4]
- A real-time fuel consumption and pollution indicators, [5]
- Surveillance applications such as proposed in [2] where nodes make videos of the road and detect and save the registration plates of vehicles around.

2. Introduction about Vehicular Sensor Networks

The VSNs are the new type of vehicular networks, whose purpose is that the real-time data collection and diffusion of information. In this the author used a VSN for a better understanding of the traffic signalling. They pointed out that the fact is vehicular sensor networks are one of the least cost solutions which tends to reduce traffic jams, CO2 emissions and fuel consumption. The proposed algorithm uses the VSNs for security issues where agent nodes can look for a stolen car for example, by sending a query to all nodes that have crossed that vehicle. Another application of VSNs is the one proposed in where the network provides the road users a more safety driving by disseminating alert messages in case of an emergency (Gurupreet singh). A VSN can be considered as a fusion of a Vehicle ad hoc network (VANET) and wireless sensor network (WSN).

The VSN has the following some properties like:

(i) Higher capacity because of the inboard sensors is supplied with the more energy, storage and computing capabilities.

(ii) High amounts of data since a vehicle could be modelled by a lot of sensors like cameras.

(iii) The management of Dynamic data will be sink since data sinks could be mobile compared with the traditional WSNs, and

(iv) Large scale connectivity is needed because wide roads and grand avenues in urban environments may contain thousands of vehicles.

These are the specific characteristics are important implications for making decisions in this networks. Thus, numerous research challenges need to be addressed for vehicular sensor networks to be widely deployed.

A. VSN Architectures

The data dissemination in vehicular sensors networks can be described based on three architectures as shown in figure 2:

- ***V2V***: In this both the collection of data and the restoring of information are done. For example this solution may be used for quick alert messages dissemination.
- ***V2I***: infrastructure based wireless links (GSM, UMTS, WiMax, Wifi/Mesh, etc.) are used to gather the data from VSN nodes.
- ***Hybrid***: The combination of both the V2V and V2I architectures.

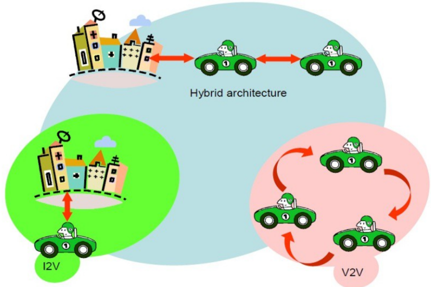

Figure 2- Vehicular communication architectures.

B. Data Dissemination in VSN

We found in the literature different approaches for data dissemination in a VSN:

1) *V2I/I2V Dissemination*

a. *Pull based dissemination*

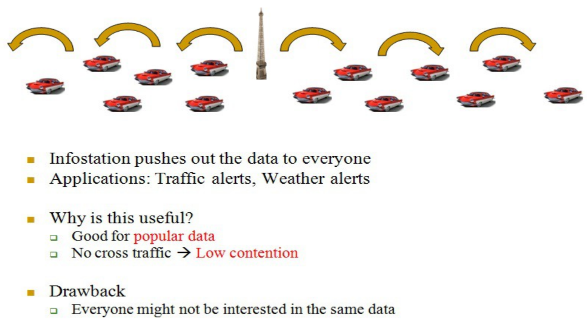

Fig 3 Pull based dissemination

b. *Push based dissemination*

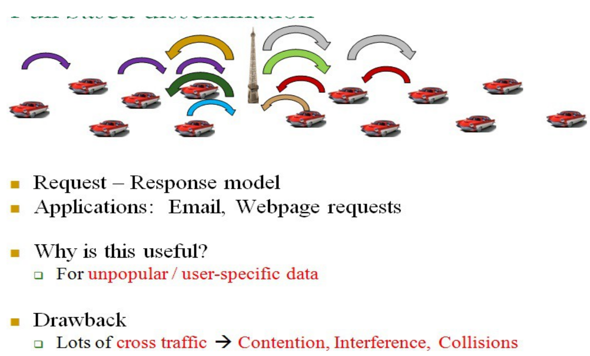

Fig 4 Push based Dessemination

In the essential of network partitioning of VSNs, it is recommended that the use of opportunistic diffusion of data, in which messages are stored in each intermediate node and forwarded to every neighbour node till the destination is reached. Thus the delivery ratio is improved. However, these types of mechanisms are not suitable for non-tolerant delay applications. Opportunistic dissemination protocols have the potential applications in the vehicular networking, ranging from advertising to emergency/traffic/parking information spreading: one of the characteristics of vehicular networks is that they are often partitioned due to lack of continuity in connectivity among cars or limited coverage of infestations in remote areas (Gurupreet Singh). Most available opportunistic, or delay tolerant, networking protocols, however, fail to take into account the peculiarities of vehicular networks.

2) *Geographical dissemination*

The fact is that the end to end paths are not present in a VSN constantly, a geographic dissemination is used in [2] by sending the message to the closest node towards the destination till it reaches it. Another way to do geographic dissemination is given in [6] where the authors show how to use geo-casting to deliver messages to several nodes in a geographical area.

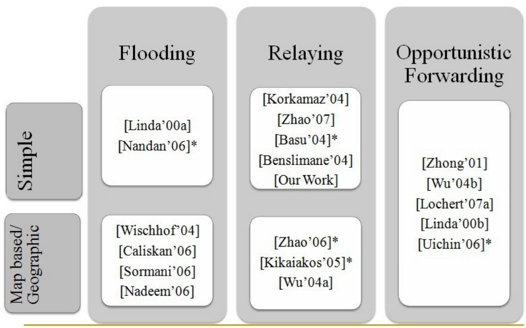

Fig 5 Geographical Dissemination

3) *Peer-to-peer dissemination*

In P2P dissemination solution, the source node stores the data in its storage device and dont send in the network, if others need it will forward. In [2], such a architecture is proposed for delay tolerant applications. The following are few peer-to-peer (P2P) systems such as BitTorrent [7], Slurpie [29], SplitStream [5], Bullet' [18] and Avalanche. The key idea is that the file is divided into M parts of equal sizes and that a user can download any one of these either from the server or from a peer who has previously downloaded it. That is, the end users collaborate by forming a P2P network of peers, so they can download from one another as well as from the server.

4) Cluster-based dissemination

In the delivery ratio and the reduction of broadcast disambiguation, a message will be relay by a minimum number of intermediate nodes to the destination. To do like, nodes are grouped as a set of clusters, in which one or more nodes (Cluster Head) gathered as data in his cluster and send them to the next cluster. Cluster-based solutions provide less propagation delay and high delivery ratio with also bandwidth equity. In [4] the authors use a distributed clustering algorithm to create a virtual backbone that allows only some nodes to broadcast messages and thus, to reduce significantly broadcast storms.

We are interested in this paper, the cluster based dissemination mechanisms are combined with the geographical structures like Trees.

C. Data Aggregation

Data aggregation is a most known concept in Wireless Sensors Networks; it allows that the concept of merging of nodes, update or delete some information because of they might be replicated, similar or expired. There are many aggregation mechanisms proposed in the sensors networks that can also be used in VSNs: In [8], a timestamp aggregation technique is developed upon an opportunistic dissemination solution. In this case, if a node receives an information, it can decide if it is valid or not, by checking its sending time. Authors of [9] use a ratio-based and a cost-based algorithm to choose which information is important to aggregate and to estimate the error that can introduce a message in the data.

3. PROPOSED EFFICIENT PROTOCOL

As in the above-mentioned section, the energy efficiency in tree-based protocol like TREEPSI is better than cluster based and chain-based protocol. If some sensor nodes send data to the sink, this information of nodes will make a detour. Thus, that will cause more power dissipation in data gathering. This situation is happened as building the binary tree paths, especially when the sensor field is large and the numbers of sensor nodes are large. In order to improve the reduction of power dissipation, we propose a novel protocol to combine the cluster-based and tree-based protocol to improve it. In the following, we will describe the deployment and method of the protocol. And the first we can see the flow chart of protocol clearly as Figure 3. According to reference above-mentioned routing protocols, the network assumptions can be initiated as follows [4, 5, and 6].

1. Each node or sink has ability to transmit message to any other node and sink directly.
2. Each sensor node has radio power control node can tune the magnitude according to the transmission distance.
3. Each sensor node has the same initial power in WSNs.
4. Each sensor node has location information.
5. Every sensor nodes are fixed after they were deployed.
6. WSNs would not be maintained by humans.
7. Every sensor nodes have the same process and communication ability in WSNs, and they play the same role.
8. Wireless sensor nodes are deployed densely and randomly in sensor field.

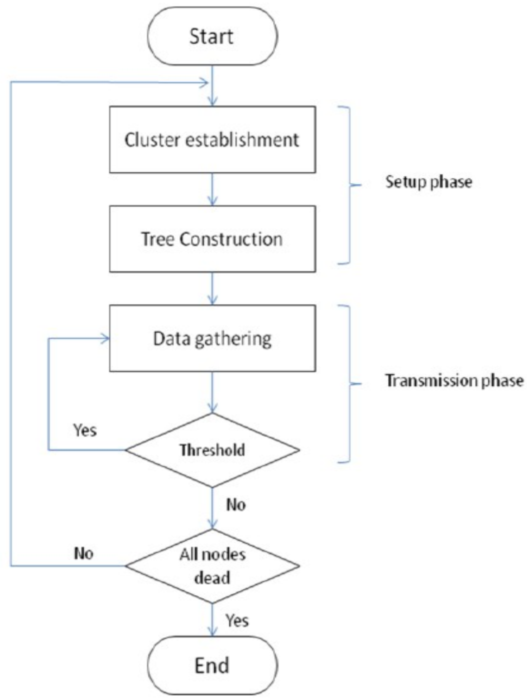

Fig 6 Flow chart for TCDGP

Sink could get the whole location and energy information about sensor nodes by two or other manners. One is recorded

in the sink at the initial state as nodes were deployed. The other is that sink broadcast whole network, and then received the back message form sensor nodes.

A. Cluster Establishment

Setup Phase:

This setup phase consists of two major steps:

1. Cluster formation and 2. Cluster head selection. Once the base station forms the primary cluster, they will not change much because of all sensor nodes are fixed, whereas the selected cluster head in the same cluster may be different in each round. During the first round, the base station first splits the network into two sub clusters, and proceeds further by splitting the sub clusters into smaller clusters.

Neighbor id	Residual Energy	Distance	Distance to BS	State	Weight

The process of cluster splitting will be repeated by base station until the desired number of clusters is attained. When the splitting algorithm is completed, the base station will select a cluster head for each cluster according to the location information of the nodes (Gurupreet Singh).

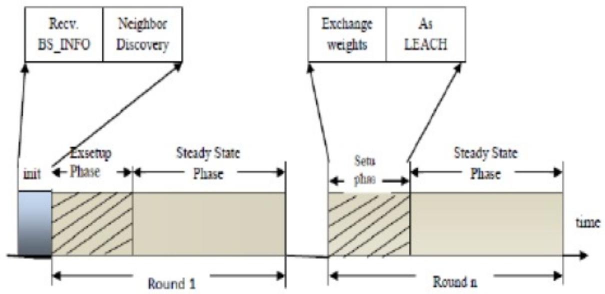

Figure-7: Extended Round Of Table

The cluster head is located at the centre of a cluster. Once cluster head is selected it broadcasts a message in the network and invites the other nodes to join in cluster with it. The other nodes will choose their own cluster heads and sends the joining message according to the power of the received broadcast messages. When the cluster head receives the joining message from its neighbour nodes, it assigns a time slot to each node to transmit data. When the first round is over and the primary cluster topology is formed, then the base station is not responsible for selecting the cluster head. The making of cluster is moved from the base station to the sensor nodes. The decision of making a new cluster head in locally in each cluster is based on the node's weight value. The pseudo code for all operation is given below:

Initialize {

1. Base station: acquire the number of clusters N;
2. Split the network into N clusters;
3. Choose cluster head from each cluster;
4. Notify the node to be cluster head.

}

Repeat:

{

1. Node i: if (Receive the notify message from the base station)
2. Work in cluster head mode;
3. If (Receive the broadcast message from cluster head node)
4. Work in sensing mode.

}

For cluster head i:

{

1. Receive data born cluster member j;
2. Compute the weight value W_i and W_j;
3. If ($W_i > W_j$), W_i Work in cluster head;
4. Else i work in sensing mode;
5. Notify j to be cluster head ;

}

B. Constructing Cluster Based Tree

Sink will collect the information that cluster head had labeled in each cluster and build path in minimum spanning tree to compute the tree path. The Minimum Spanning tree (MST) concept in the Greedy algorithms used to solve the undirected weight graph problem. After eliminating some of the connection links, the sub-graph still have the connection ability. For this reason, sub-graph can reduce the sum of the weights. A sub-graph who has the minimum sum of weights must be a tree like framework. Spanning tree could let all nodes conform to tree definition which is connected in the graph. A connected sub-graph which has a minimum sum of weights must be a spanning tree. On the contrary, it is not correctly absolutely. There could be several kinds Minimum spanning tree in a graph, and it is not the only one. But their sum of weight should be the same. If we use Brute Force to find the minimum spanning tree, it will produce huge computation time. In order to avoid this, we use Prim algorithm to help us finding the MST.

C. Data Aggregation

After the routing mechanism has established, every tip nodes transmit gathering data to upper level nodes. Then the upper level nodes will fuse received data and sensed data by itself, and send the data to next upper level nodes. The process will keep going until the root node, cluster head, has aggregated the data in the cluster. It is called a round‖ as all root nodes has finished transmitting data.

4. PERFORMANCES EVALUATION

To validate and evaluate TCDGP, we have chosen Qualnet 4.5 simulation environment. We also extended and adapted the mobility model proposed in [11] to our needs. Our tool generates realistic random vehicles' displacements.

A. Assumptions

1) Spatiotemporal environment

We execute TCDGP on a straight road section partitioned into 18 equal segments, as depicted in Figure 9. The base station that covers all the section is present at one end point of the road. All the key parameters of our simulation are summarized in the following table:

SIMULATION / SCENARIO		MAC / CGP	
Simulation time	600s	MAC protocol	802.11b
Map size	2500x2500 m2	Capacity	2 Mbps
Mobility model	VanetMobisim [12]	Trans. Range	~266 m
Number of seg.	18	HL_PT	~0.1 s
Nodes	50 - 1000	PK_PT	~0.2 s
Vehicle velocity	0 – 108 km/h	IS_PT	~0.1 s
Segment length	100 m	FULL _DURATION	5 s
Road length	1.8 km	CHD_DURATION	1 s
Road width	15 m	GAT_DURATION	3 s
Number of lanes	2	AGG_DURATION	~0.1 s
Store, carry and forward	Not used	DIS_DURATION	1 s

Table 1: Simulation Setup

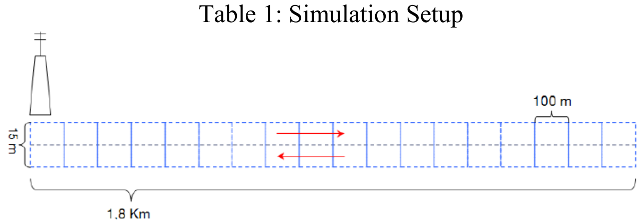

Fig.8 Spatial Environment

B. Simulation Scenarios

1) Scenario 1: Per node dissemination

In this scenario, each node sends its collected data (speed, position, etc.) individually and periodically to the base station using the provider's cellular network. The aggregation in this case, is done at the Telco provider level. (See Figure 9)

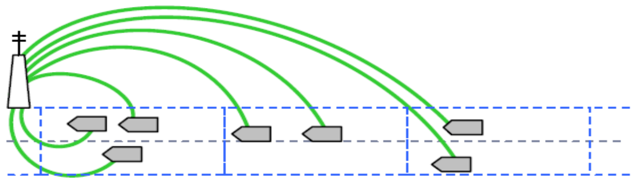

Figure 9 – Per node dissemination scenario.

2) Scenario 2: Per Cluster Head dissemination

In this scenario (see Figure 10), the local data gathering and aggregation are done at the segment level, as described in TCDGP. The aggregated data (average speed, number of nodes, etc.) are sent to the base station directly from the cluster head of each segment. The Telco provider will only aggregate the data from each segment.

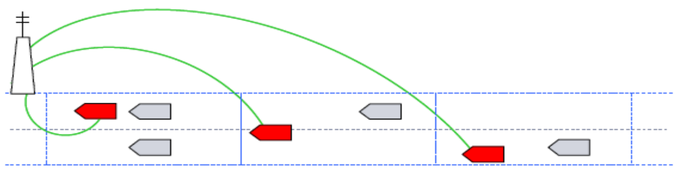

Figure 10 – Per cluster head dissemination scenario.

3) Scenario 3: Complete TCDGP dissemination

As depicted in Figure 8, TCDGP will be integrally executed in this scenario, from the cluster head election to the data dissemination to the provider.

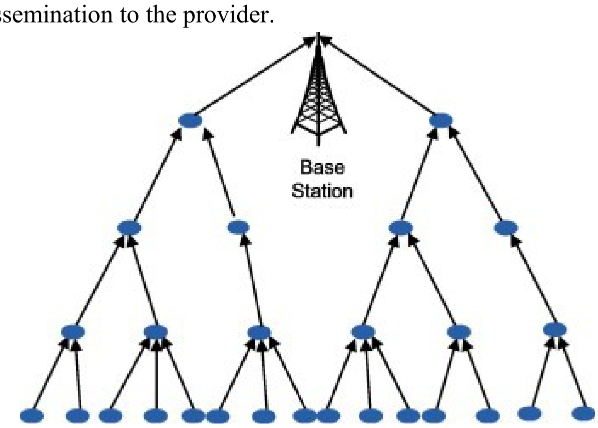

Figure 11- Complete TCDGP dissemination

C. Simulation Results

We calculate the number of messages sent to the base station via the provider's cellular network during 600 seconds.

Thus, we can see in which scenario the data collection is the greediest in terms of cellular network usability.

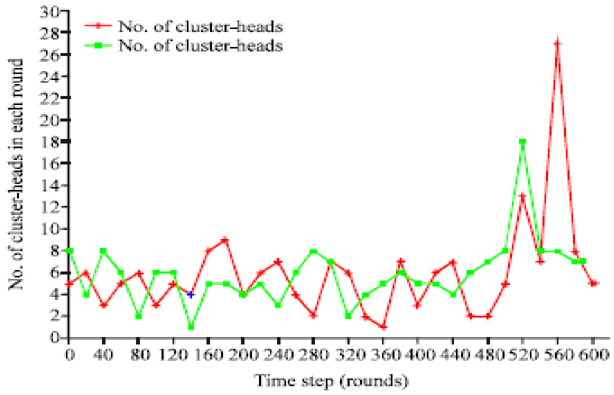

Figure 12 – Numbers of V2I messages

5. CONCLUSION

In this paper, a novel data gathering and dissemination system (TCDGP, Tree based Clustered Data Gathering Protocol) based on hierarchical and geographical dissemination mechanisms on vehicular sensors networks is proposed. Designed for hybrid VANET architecture, it allows telecommunication/service providers to get valuable information about the road environment in a specific geographical area, using V2V network to minimize the high-cost links usability and base stations to gather information from the vehicles. Simulations results of TCDGP demonstrate the feasibility of the proposed approach; moreover, they show that TCDGP reduces considerably the provider's network usability without any loss of accuracy in the collected data. We are currently extending this work by performing other extensive simulation in order to study all the TCDGP parameters.

REFERENCES

[1] Ismail Salhi, Mohamed Oussama Cherif, and Sidi Mohammed Senouci: A new Framework data collection in vehicular sensor networks. 2009.

[2] C. Intanagonwiwat, R. Govindan, and D. Estrin. "Directed Diffusion: A Scalable and Robust Communication Paradigm for Sensor Networks." ACM MobiCom'00, August 2000, Boston, Massachusetts.

[3] U. Lee, E. Magistretti, B. Zhou, M. Gerla, P. Bellavista, A. Corradi, "Efficient Data Harvesting in Mobile Sensor Platforms". IEEE PerSeNS'06 Workshop, March 2006, Pisa, Italy.

[4] R. Panayappan, J. M. Trivedi, A. Studer, A. Perrig, "VANET-based Approach for Parking Space Availability", ACM 4th workshop on VANETs, September 2007, Montreal, Quebec, Canada.

[5] L. Bononi,M. Di Felice, "A Cross Layered MAC and Clustering Scheme for Efficient Broadcast in VANETs", IEEE MASS'07, October 2007, Pisa, Italy.

[6] E. Guizzo. "Network of traffic spies built into cars in Atlanta". IEEE Spectrum, April 2004.

[7] H. Wu , R. Fujimoto , R. Guensler , M. Hunter, "MDDV: a mobilitycentric data dissemination algorithm for vehicular networks", ACM international workshop on VANETs, October 2004, Philadelphia, , USA.

[8] U. Lee, E. Magistretti, M. Gerla, P. Bellavista, A. Corradi, "Dissemination and Harvesting of Urban Data using Vehicular Sensing Platforms", submitted for publication.

[9] L. Wischhof, A. Ebner, H. Rohling M. Lott and R. Halfmann. "SOTIS - A Self-organizing Traffic Information System". IEEE Vehicular Technology Conference, April 2003.

[10] T. Nadeem, S. Dashtinezhad, C. Liao and L. Iftode. "TrafficView: Traffic Data Dissemination using Car-to-Car Communication". ACM SIGMOBILE'04, July 2004.

[11] C. Lochert, B. Scheuermann and M. Mauve. "Probabilistic Aggregation for Data Dissemination in VANETs", ACM VANET'07, September 2007, Montreal, Quebec, Canada.

[12] P. Flajolet and G. N. Martin. "Probabilistic counting algorithms for data base applications". Journal of Computer and System Sciences, September 1985, Orlando, FL, USA.

[13] J. Haerri, M. Fiore, F. Filali, C. Bonnet, “VanetMobiSim: generating ealistic mobility patterns for VANETs”, ACM VANET’06, Sept. 2006.